L'activité de méta-évaluation des politiques publiques et des programmes associés et le besoin de qualité des connaissances issues des évaluations.

Professeur Maurice Baslé.
CEREQ-CREM-CNRS-Université Rennes 1.
Chaire Connaissance et action territoriale Fondation Université de Bretagne sud.
18 Avril 2013. Paris.

Séminaire Sfe du réseau des chercheurs en Evaluation des politiques publiques.
Document en cours de publication. Tous droits réservés.

Questions.

La méta-évaluation est souvent simplement définie comme l'évaluation d'une ou des évaluations (évaluation niveau 2). Peut-on affiner le concept ? Peut-on ensuite décrire l'activité de Méta-Evaluation ? La question posée ici est la suivante : un niveau deux d'évaluation doit-il être organisé pour rendre capitalisable les évaluations ? On montre que les synthèses existent et se font sur la base de critères de qualité des évaluations. Ex post, elles ne sont pas toujours possibles. Mais on peut aussi construire un projet ex ante de futures méta-évaluations. Peut-on faire un constat de qualité à partir de la valeur ajoutée ou de l'utilité obtenue par la méta-évaluation choisie ?

Résumé.

Quatre approches et un constat.

1. **La méta-Evaluation peut être définie comme l'audit ex post des forces et des faiblesses d'UNE évaluation**.
   Il a été commode de commencer par définir la méta-évaluation comme une activité d'évaluation d'une évaluation. C'est alors une espèce de niveau 2 de l'évaluation. C'est ce qui a été proposé dans la littérature depuis longtemps. Voir par exemple : *"Meta-evaluation is the evaluation of evaluation"* (Cook & Gruder, 1978; Scriven, 1969) in : Shadish W.-R., 1998, Evaluation Theory is Who We Are, *American Journal of Evaluation*, Vol. 19 , N°1, pp 1-10.
2. **La méta-évaluation peut être définie comme une synthèse ex post DES évaluations se rapportant à une même politique ou à ses programmes.** Il y a bien une évaluation de plusieurs évaluations, l'activité étant demandée ex-post. C'est une forme de compilation de ce qui a été produit dans différentes évaluations déjà conduites. On en espère une « super-connaissance » et un bilan vraiment complet. Ce bilan peut être une capitalisation des résultats utilisables, une progression de la connaissance évaluative utilisable. **La méta-évaluation de synthèse ex post** fait appel à des super-critères de qualité au regard desquels il faut évaluer cette fois de manière systématique PLUSIEURS évaluations précédemment effectuées[1].

[1] Metaevaluation is an assessment of an evaluation, especially its adherence to pertinent standards of sound evaluation (Stufflebeam Daniel. *Program Evaluations Metaevaluation Checklist.* www.wmich.edu/evalctr/checklists)

3. **La méta-évaluation peut-être aussi définie comme un projet ex ante de mise en place des évaluations pour des programmes identiques mais déployés dans le temps ou dans différents contextes.** C'est un projet approfondi de supervision d'évaluations tentant de produire à plusieurs reprises la même évaluation de la même politique et d'en retirer plus de connaissances de ce qui « fonctionne » ou pas dans les processus et les chaines logiques « réalisations- résultats et impacts ». La force de la preuve de l'existence d'un « mécanisme » ou « processus » qui marche ou pas doit en être renforcée. C'est la tache des « knowledge capitalisation units » de capitaliser.[2]
4. **Encore plus sur le mode projet ex ante, on peut aussi imaginer en variante du déploiement précédent, la méta-évaluation comme une supervision d'expérimentations parallèles rendues ex ante évaluables dans les mêmes termes.** Modulo les limites dues à la construction des conditions expérimentales (« artificielles ») en vue d'une évaluation d'expérimentations, il doit alors être possible de tester la validité externe de résultats concernant les processus qui marchent et les processus qui ne marchent pas. D'autres expérimentations peuvent à l'issue d'une première méta-évaluation et méta-analyse être mises en place.

Mots-clés. Le référentiel et les standards. La check-list de la Méta-Evaluation. Qualité totale. Les super-critères de méta-Evaluation. Méta-Evaluation, apprentissage et capitalisation des connaissances évaluatives.

Méta-évaluation

Meta-evaluation

Evaluation d'une autre évaluation ou d'une série d'autres évaluations.

Les critères de jugement sont généralement la fiabilité, la crédibilité et l'utilité. Le terme est quelquefois appliqué à l'audit de la fonction évaluative d'une organisation (vérification que les règles concernant l'évaluation ont été appliquées avec professionnalisme). On parle parfois de méta-évaluation pour une synthèse faite à partir d'une série d'évaluations.

Means. 1999

[2] Comme à l'AFD. http://www.afd.fr/lang/en/home/recherche/evaluation-capitalisation/evaluation-et-capitalisation

Trois remarques préalables doivent être faites. D'une part, la littérature n'est pas centrée sur méta-évaluation des politiques publiques mais sur la méta-évaluation. C'est une approche souvent très englobante de la réflexivité sur les pratiques d'évaluation, en prévoyant ex ante, en inventant ex post ou en proposant chemin faisant de faire une méta – évaluation chemin faisant[3]. D'autre part, la position réflexive de la méta-évaluation est évidemment celle de l'activité de Recherche en SHS et aussi celle de l'évaluation de cette recherche en SHS. Enfin, les questions d'arbitrage entre ressources consacrées aux évaluations et budget consacré à la méta-évaluation ne sont pas souvent évoqués.

Quatre approches du concept et de la pratique de la méta-évaluation se trouvent dans la littérature compilée ici. On distingue selon qu'une **évaluation ou plusieurs** sont passées en revue ou analysées ou synthétisées. On distingue également **selon le temps** de la méta-évaluation (conçue ex ante, déployée volontairement pendant, inventée ex post …). Dans une acception particulière, la méta-évaluation s'appuie sur le déploiement conçu à l'avance **d'expérimentations en parallèle** : dans ce cas, les possibilités de la méta-analyse[4] rejoignent ceux que l'on trouve dans certains cas dans l'activité de méta-évaluation.

Dans tous les cas, la définition consiste à inventer et concevoir la méta-évaluation pour renforcer les connaissances évaluatives de toutes natures en prenant un point de vue réflexif (interne ou/et externalisé) en vue d'une meilleure compréhension de la confiance que l'on peut accorder aux résultats produits et de leur « caractère utilisable » dans les politiques publiques à venir. En ce sens, l'activité de méta-évaluation correspond à un mode d'apprentissage permanent : apprentissage organisationnel, apprentissage dans la production et le management des connaissances, apprentissage des politiques[5]. L'activité combine des expertises, des approches et multiplie les points de vue, ce qui en fait une activité transdisciplinaire.

1. **Méta-évaluation comme évaluation ex post et non prévue d'une évaluation primaire.** Au niveau 1, un cahier des charges a été lancé. L'évaluation se déroule sur la base de ce cahier des charges. Elle suscite de l'espoir d'en savoir plus. Elle choisit sa gouvernance, ses alliances pour retrouver des informations, sa manière de faire en préconisant tel ou tel type d'outils, elle est plutôt managériale, plutôt pluraliste, plutôt centrée sur les stratégies d'acteurs ou la mise en œuvre, plutôt centrée sur le suivi des indicateurs de résultats… Elle rencontre habituellement des difficultés et doit souvent s'adapter au terrain[6]. Il est nécessaire de faire le point. Il apparait que

[3] Voir par exemple Hanssen, C.E. Jawrenz, F. Dunet, D.O. (2008). Concurrent Meta-Evaluation. *American Journal of Evaluation.* Vol. 29, N°4. December. Pp.572-582.

[4] La méta-analyse renvoie à une analyse statistique des analyses en vue d'intégrer des résultats.

[5] Baslé, M. 2003. Les gouvernements apprennent-ils. Les Echos.
http://www.pnud.ne/rense/Biblioth%E8que/Basl%E902.pdf

[6] "These classes are conceptual, sociopolitical, contractual/legal, technical, administrative, and moral/ethical. Basically, these problems are suggested for use in improving ongoing evaluation work (the matter of formative meta-

l'on est loin des objectifs de l'évaluation et qu'on doit au minimum réfléchir sommairement au degré de confiance que l'on peut attribuer aux résultats de l'évaluation. La démarche d' »inventaire » est lancée. C'est un retour sur le mode de production des résultats, ce retour est souvent le texte du contre rapporteur de l'évaluation. Pour que l'on puisse parler de méta-évaluation, il ne suffit pas de ce referee rapide, on conviendra que la démarche doit être à nouveau codée comme une nouvelle activité et qu'un nouveau cahier des charges doit être lancé qui doit partir des différentes dimensions de « qualité » et des « forces et faiblesses du processus précédent d'évaluation et de ses résultats ». Ainsi, selon Stufflebeam, la Métaevaluation est un processus pour auditer les forces et les faiblesses d'UNE evaluation : *"the process of delineating, obtaining, and applying descriptive information and judgmental information about an evaluation's utility, feasibility, propriety, and accuracy and its systematic nature, competence, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and publicly report its strengths and weaknesses".*[7]

La méta-évaluation au sens d'évaluation d'une évaluation est vraiment l'appréciation de la qualité d'une évaluation. Certains parlent de référence à un référentiel, à un « professionnalisme », à une « grande expérience ». **Scriven propose qu'un « evaluation advisor» soit associé en amont à chaque évaluation,** de telle sorte que l'on réduise le besoin de revenir ensuite à la méta –évaluation ainsi conçue[8].

Déjà, à ce niveau simple d'évaluation d'une évaluation, disons que la qualité s'examine dans différentes dimensions qui sont celles des standards de l'évaluation. On peut ainsi se référer aux cinq critères du petit guide de l'évaluation du CSE[9]. Pour Guba et Stufflebeam, it « *must produce findings that are internally and externally valid; i.e., the findings must be true and they must be generalizable. While evaluations must also meet these technical standards, they are not sufficient for judging evaluation findings. In addition to producing good information; i.e.technically sound information, evaluation must produce findings that are useful to some audience; and the findings must be worth more to the audiences than the cost of obtaining the information--the concern of cost/effectiveness*"[10]. Pour Scriven, la

---

evaluation), but they should also prove useful in assessing and diagnosing completed evaluation studies (a concern of summative metaevaluation)"Stufflebeam. 1974.

[7] *Source :* 'Stufflebeam Daniel L., 2001, « The Metaevaluation Imperative », *American Journal of Evaluation*, Vol. 22, No. 2, pp. 183–209. Les travaux se rapportent au domaine de l'Education.

[8] http://michaelscriven.info/images/EVALUATION_ADVISOR.6.3.11.pdf

It is a person who serves as a helper or guide about evaluation, but not as an evaluator

[9] Conseil scientifique de l'évaluation (1996). Le petit guide de l'évaluation. La documentation française. Les cinq critères ne sont en réalité pas tous de même nature : utilité-pertinence, fiabilité, objectrvité, généralisabilé, transparence. Ils comprennent des facteurs d'attitude et des résultats d'analyse.

[10] These three standards of technical adequacy, utility and cost/effectiveness have been spelled out by Guba and Stufflebeam (1970).

check-list de la qualité est la suivante : critères de qualité du cahier des charges, d'éthique ; existence d'un recoupement des résultats…. débouchant sur un degré de confiance dans les résultats, et éventuellement la possibilité de généraliser les résultats[11].

Le travail de méta évaluation d'une évaluation peut être prévu en « chemin faisant ». On parle alors deb « concurrent meta-evaluation ».

Cet examen ne se fait ici que pour une SEULE évaluation, il est possible d'envisager le même processus de supervision analytique et critique pour PLUSIEURS évaluations.

2. **Méta-Evaluation comme évaluation potentiellement utile DES évaluations en vue de l'élaboration ex post d'un tableau de synthèse selon des dimensions de résultats.**

L'évaluation de grands projets est quelquefois appelée « méta-évaluation » en raison de la grande taille du programme évalué (évaluation du méta !) ou de l'évaluation de la somme des projets se rapportant au même grand évènement (exemple : le cas des jeux olympiques de Londres en 2012) .

Plus généralement, quand plusieurs évaluations d'une même politique ont été réalisées, les évaluations sont généralement de différents types avec différents cahiers des charges.

Côté dimension d'objectifs, certaines peuvent aborder des dimensions d'objectifs et de résultats plus particulièrement que d'autres.

Côté outillage (méthodes, moyens de documentation…), certaines font appel à des sources spécifiques, d'autres créent des nouveaux documents qui leur sont propres. Ce rapport à la connaissance évaluative n'est pas neutre, les alliances des différents évaluateurs avec les « sachant » du programme peuvent en effet être différentes.

Lorsqu'on assemble les résultats des évaluations et qu'on cherche à croiser, comparer, il apparait que les connaissances recherchées ne sont pas homogènes et conciliables. En général, elles ont plusieurs facettes elles mêmes issues de plusieurs types de mélanges de données brutes croisées et retraitées. Chaque instance, les services, les experts prestataires produisent une « connaissance évaluative « dépendante du contexte. Il n'est donc pas toujours aisé in fine de synthétiser les

[11] http://michaelscriven.info/images/EvaluatingEvals-Checklist.pdf
La nécessité de la check-list a été discutée. Cf Final report London olympic games. "Whilst this ambition has not gone away, expectations for the utility of the checklist approach (especially) have waned. As evaluators employed a wider and wider set of approaches and tackled a wider and wider range of interventions the prospects diminished of subsuming all requirements under a simple, useful set of guidelines. The result was that the listing of standards multiplied. Crucially, in attempting to cover the diversity of modern evaluation practice, the standards become more and more abstract (to the point perhaps of losing all utility").

connaissances que l'on trouve dans les différentes facettes des livrables, voir ou tenter de voir si elles sont ou non contradictoires par rapport à des réponses à des questions évaluatives.

L'activité de méta-évaluation est basée sur la croyance que c'est possible de se donner une grille de lecture des cahiers des charges avec des critères ou des dimensions d'intérêt et que ceci va permettre des regroupements et une « synthèse »[12].

L'activité de méta-évaluation consiste à passer en revue systématiquement les résultats DES évaluations au regard de super-critères de qualité des savoirs produits. Ces super-critères sont finalement assez proches de ce que l'on trouve dans les guides d'évaluation et dans ce que les sociétés professionnelles d'évaluateurs ont posé comme standards et déontologie pour les métiers des évaluateurs[13]. Si ces super-critères n'existaient pas, il n'y aurait pas de méta-évaluation, tout juste un regard sur quelques qualités de chacune des évaluations. Dès lors qu'ils existent, ils peuvent permettre d'avoir une approche synthétique et de « noter » les qualités et défauts DES évaluations.

Plus profondément, à partir des regards sur la qualité des évaluations de la même politique tant au plan de la forme qu'au plan du fond, on suppose la synthèse possible en organisant la connaissance évaluative et les retours de savoirs de différentes approches évaluatives.

Dans le cas positif, la méta-évaluation produit 1°) une synthèse des résultats existants sur telle ou telle politique vis-à-vis de populations ou de territoires d'application, cette synthèse pouvant donner lieu à un tableau de synthèse. La synthèse peut faire apparaitre des retours de savoirs évaluatifs différents selon les angles d'attaque et les cahiers des charges. 2°) de véritables capitalisations de savoirs sur les politiques concernées : l'activité de méta-évaluation permet alors, grâce à une analyse et une connaissance produite plus approfondie des mécanismes de succès, des recommandations plus assurées et plus adaptées pour améliorer ou réformer les politiques publiques.

Un tableau de synthèse peut être constitué au regard d'une check-list des dimensions de qualité des évaluations, par exemple, pour Stufflebeam :

[12] Cf. London Olympic games final report. "Meta-evaluations can also be seen as a form of systematic review where the findings of individual studies are combined (synthesised) to answer an evaluation research question.

[13] Voir charte de la société française d'évaluation par exemple ».

Conseil scientifique de l'évaluation. 1996. Le petit guide de l'évaluation. La Documentation française.

"Program Documentation, Context Analysis, Defensible Information Sources, Valid Information, Reliable Information, Systematic Information, Analysis of Quantitative Information, Analysis of Qualitative Information, Justified conclusionsImpartial Reporting[14].

Il existe des cas où les connaissances de synthèse n'existent pas. On s'en tient à une « collection de cas » ou à un rassemblement de connaissances par facette. Ce peut être tout particulièrement observé dans le cas de politiques complexes[15].

Exemple.

> ***"The Swiss Innovation Agency (CTI**) has administrated its collaborative research funding scheme since the early 1980s. Between 1989 and 2002 the scheme was evaluated 14 times. In a study combining meta-evaluation and evaluation synthesis, we assessed these existing evaluations against a variety of criteria, including selected evaluation standards of the Swiss Evaluation Society. The meta-evaluation showed that the existing evaluations were mostly qualitative, internal and ex post and that evaluation practice at the CTI was episodic. The evaluations differed in quality, with most evaluation standards being fulfilled adequately to fairly well. The results of the meta-evaluation were central to the ensuing evaluation synthesis by giving information on the adequacy of the existing evaluations. The synthesis compiled the—mostly qualitative—results of the evaluations. Our study shows that there are patterns of outputs and outcomes in evidence across the evaluation studies that suggest that CTI funding is having an effect in terms of outputs and outcomes. But it also shows that there are gaps in the evidence base of the evaluation studies that make it difficult to come to more precise conclusions"[16].*

3. La Méta-évaluation comme déploiement ex ante de la même politique à évaluer dans différents contextes.

La méta-évaluation peut être prévue à l'avance lorsque l'on programme la conduite du déploiement de programmes parallèles et rendus évaluables dans les mêmes termes. Il est alors prévu de produire des données brutes qui peuvent être ensuite recoupées et

---

[14] **PROGRAM EVALUATION MODELS METAEVALUATION CHECKLIST, (Based on *The Program Evaluation Standards*), Daniel L. Stufflebeam, 1999**

[15] Sfe publications. *Evaluer et démêler la complexité des politiques publiques.* Avril 2013.

[16] **Assessing the effects of a collaborative research funding scheme: An approach combining meta-evaluation and evaluation synthesis**

1. Barbara Good*

+Author Affiliations

1. Technopolis Group Austria, Rudolfsplatz 12/11, 1010 Vienna, Austria

↵*Corresponding author. Email: barbara.good@technopolis-group.com

analysées et il semble possible de tester la validité externe de résultats concernant les processus qui marchent et les processus qui ne marchent pas.

Exemples.

1. **Unicef. Research Methods and Constraints**

UNICEF. … "3 The meta-evaluation analyzes 59 CP evaluation reports that were produced between 2002 and 2007, or 40% of all CP evaluations in that period. Selection was based upon availability and language (English). Scales were established to assign values to characteristics of interest addressing programming quality, evidence quality, and PEF outputs and outcomes. 4. Each report's catalogue of PEF investments was coded for delivery into the askSam database, which allows text searching of materials from a variety of document types. 5. Limitations: in most cases, the depth of the database for identifying best practices is constrained by (a) insufficient analysis of institutional contexts (e.g. donor requirements, host government sensitivities), (b) scarce time series information (e.g. 64% of the reports have no longitudinal data), and (c) an alarming over-reliance on Key Informant Interviews with an under-emphasis on sampling and scale (e.g. 80% lack a sampling plan)".

Autre exemple. Lynne R. Tingle, Marie DeSimone,Benjamin Covington A Meta-Evaluation of 11 School-Based Smoking Prevention Programs. Article first published online: 9 OCT 2009, DOI: 10.1111/j.1746-1561.2003.tb03574.x. 2003 American School Health Association

4. **Variante. Méta-évaluation et déploiement expérimental de la même politique dans différents contextes. Cette fois, le projet méta-évaluatif est conçu d'emblée comme un processus de mise en œuvre d'expériences placées autant que possible sous contrôle. Une méta-analyse peut être l'issue.**

   Le contrôle du déploiement d'un programme peut être mis en place comme le déploiement d'expérimentations en parallèle dans différents contextes.

   A la sortie d'un tel processus, les résultats peuvent faire l'objet d'une méta-analyse. Le degré de confiance dans les résultats qui en résultent n'est cependant pas encore total car, généralement, le processus expérimenté est à petite échelle ou petit domaine et produit des effets à un « bas niveau » qui doivent être interprétés à ce niveau mais aussi à un niveau supérieur là où les effets indirects d'autres programmes peuvent s'avérer importants même dans le cas d'expériences bien menées dès l'origine.

   Le contrefactuel bien produit donne cependant de sérieuses indications pour la « facette » de la connaissance que l'on a cherché à produite, même si le besoin d'avoir des résultats sur plusieurs facettes demeure[17].

   Exemple. ***"****Among the initiatives included in the National ITS Program Plan is the conduct of operational tests related to the various user services. These tests are conducted for a prototype system on a scale smaller than full deployment over a relatively short period of time in a "real-world" (as opposed to a controlled laboratory*

[17] Agnès Benassy-Quéré. Evaluation des politiques publiques. *Les Notes du Conseil d'analyse économique*. CAE 2013. 1.

or otherwise contrived) environment. One purpose of the tests is the evaluation of systems of ITS technologies that are wholly or in part beyond the R&D stage, but not yet ready for full deployment. Another purpose is the evaluation of user service benefits. Such evaluations will help to identify the more promising services and technologies (in terms of their impacts on ITS program
objectives) for further development and deployment.
Multiple operational tests may be conducted for a particular user service. These tests are evaluated to provide one or more outcome measures of interest. When evaluation results for a number of operational tests become available, then it may be possible to synthesize results across tests to arrive at composite outcome measures. This synthesis process, called meta-evaluation, can provide information that is useful for determining the value of full scale deployment of an ITS user service"[18]
Voir aussi
Magura, S. (2012). Failure of intervention or failure of evaluation: A meta-evaluation of the National Youth Anti-Drug Media Campaign evaluation. Substance Use and Misuse, 47, 1414–1420. CopyrightC 2012 Informa Healthcare USA, Inc.
Laura R. Peck, Social and Economic Policy Division in Substance Use & Misuse, 47:1425–1426, 2012 : "There is no reason that we could not randomize delivery of media campaigns, for example, to learn about their effectiveness".
Need for and difficulty of 'meta-evaluations'. Greenberg, David and Mark Shroder (2004), The Digest of Social Experiments. Washington, DC: The Urban Institute Press.

5. La Méta-Evaluation comme activité pluri-disciplinaire.

Les disciplines ont, surtout en France, des codes d'entrée dans l'activité de recherche. Ceci a produit des effets de segmentation des approches évaluatives des chercheurs, enseignants-chercheurs et des consultants spécialisés (pour la majorité qui n'est pas « transdisciplinaire »). La multiplication des supports de publications a cependant permis l'éclosion de supports aux interfaces[19].

Nul ne peut échapper cependant au choix d'une entrée par un « regard » particulier. Les évaluations peuvent s'appuyer successivement ou simultanément sur plusieurs modes de de production de connaissances utiles à la compréhension de la politique publique, de sa mise en œuvre et de ses effets.

---

[18] A Data Fusion Framework For Meta-Evaluation of Intelligent Transportation System Effectiveness
William M. Evanco. 1996.

[19] Voir par exemple. Measuring Science, Capita Selecta of Current Main Issues , Anthony F.J. van Raan. In Handbook of Quantitative Science and Technology Research, The Use of Publication and Patent Statistics in Studies of S&T Systems, edited by Henk F. Moed, Centre for Science and Technology Studies, University of Leiden, The Netherlands Wolfgang Glänzel Steunpunt O&O Statistieken, K.U. Leuven, Belgium and Ulrich Schmoch Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

Jacob, S. (2008). Cross-Disciplinarization. A New talisman for Evaluation ? *American Journal of Evaluation.* Vol. 29. N°2. June. *Journal of Multidisciplinary Evaluation*. http://survey.ate.wmich.edu/jmde/index.php/jmde_1

Joseph Trimble, Ed Trickett, Celia Fisher, and Leslie Goodyear. 2012. **A Conversation on Multicultural Competence in Evaluation.** *American Journal of Evaluation* . March 2012 33: 112-123, first published on December 15, 2011.

Michael Morris. 2011. The Good, the Bad, and the Evaluator: 25 Years of AJE Ethics. *American Journal of Evaluation*. March 2011-32: 134-151, first published on November 11, 2010.

Stufflebeam, D. L. (2001). Interdisciplinary PhD programming in evaluation. *American Journal of Evaluation*, *22*, 445-455.

Conclusion.

Au final, il apparait que c'est en raison de la diversité des approches évaluatives d'une même politique (puisqu'il s'agit de « politiques publiques » qui sont par essence « controversées »), et aussi en raison de l'élasticité de la notion de référence à des standards que la Méta-Evaluation existe et a un grand avenir devant elle.

L'activité d'évaluation des politiques publiques et des programmes s'insère en outre dans un contexte nouveau de société d'information[20] et de projet de société fondée sur la connaissance[21]. Les volontés d'ouverture et de partage des données publiques publiées participent d'une volonté de gouvernance plus « ouverte »[22].

Cette capacité cognitive sera construite pas à pas sur la base de la pluridisciplinarité des approches et du recours sans complexe à différents types d'outils ad hoc.

La méta-évaluation peut avoir un développement « ordonné et pensé », l'on peut développer progressivement» des capacités de production de nouveaux savoirs évaluatifs à des fins de re-conception des politiques publiques dans un domaine particulier.

---

[20] Maurice Baslé, Thierry Pénard **(2002).** *E-Europe - La Societe europeenne de l'information En 2010.* Editions Economica. Paris.

[21] Maurice Baslé, Michel Renault (2004). *L'économie fondée sur la connaissance. Questions au projet européen.* Editions Economica. Paris. Maurice Baslé (2010). *Connaissance et action publique.* Editions Economica. Paris.

[22]En France, ce sont les objectifs du Comité interministériel pour la modernisation de l'action publique. (2013). Deux avril. http://www.fonction-publique.gouv.fr/fonction-publique/la-modernisation-de-la-fonction-publique-0

L'activité de Méta-Evaluation, si des moyens lui sont alloués, pourrait être la base d'une réponse à la question « les gouvernements apprennent-ils ? »[23]. Sisyphe ou l'éternel recommencement ou le début d'un vrai Publipédia !

[23] http://www.researchgate.net/researcher/11152001_Maurice_Basle/

Repères bibliographiques pour Méta-Evaluation.

Stufflebeam, D.L., (2001), « The Metaevaluation Imperative », *American Journal of Evaluation*, Vol. 22, No. 2, pp. 183–209.

Stufflebeam, D. P*rogram Evaluations Metaevaluation Checklist.* www.wmich.edu/evalctr/checklists

Oreskes, N. (2004), Environmental Science & Policy 7 pp.369–383. Science and public policy: what's proof got to do with it?

Hanssen, C.E. Jawrenz, F. Dunet, D.O. (2008). Concurrent Meta-Evaluation. *American Journal of Evaluation.* Vol. 29, N°4. December. Pp.572-582.

Uusikylä P. Virtanen, P. (2000). Meta-Evaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland. *Evaluation* . Sage. January 2000 6: 50-65.

Meta-Evaluation of London Olympic games. Final report. 2012.

Glass, G.V. (1976). American Educational Research Association Primary, Secondary, and Meta-Analysis of Research. . Reviewed work(s): Source: Educational Researcher, Vol. 5, No. 10 (Nov., 1976), pp. 3-8Published by: American Educational Research Association
Stable URL: http://www.jstor.org/stable/1174772 .

Annexe. Le petit guide de l'évaluation. CSE. 1996

Les cinq critères retenus par le CSE

La qualité d'une évaluation ne se résume pas à sa validité scientifique. Il faut également prendre en compte les finalités sociales de l'évaluation. C'est par rapport aux trois dimensions (dimensions cognitive, normative et instrumentale) que doit être appréciée la qualité d'une évaluation.

Utilité-pertinence : c'est le fait que l'évaluation produise de l'information compréhensible et utile, non seulement pour les décideurs mais aussi pour les autres protagonistes de la politique publique. Pour cela, le rapport d'évaluation doit répondre de manière directe et compréhensible aux questions posées dans le projet. Le critère peut également inclure le respect des délais. Enfin, est également pris en compte l'intérêt de l'apport cognitif pour la société. Les réflexions produites par l'évaluation peuvent en effet contribuer de manière diffuse à former le jugement de l'opinion publique sur la politique évaluée et à son information sur les problèmes de société directement liés à la politique. Elles peuvent également faire progresser la recherche et la réflexion politique sur les problèmes liés à la politique évaluée.

Fiabilité : l'évaluation doit être "digne de confiance" (trustworthiness). Cela renvoie à l'exactitude des informations collectées et à la valeur scientifique des imputations causales sur lesquelles se fonde le jugement évaluatif. Il convient en particulier d'être attentif aux biais que les techniques de collecte et de traitement des informations peuvent introduire dans la production des conclusions.

Objectivité : l'objectivité s'entend comme le fait que les conclusions de l'évaluation n'ont pas été influencées par les préférences personnelles ou les positions institutionnelles des responsables de l'évaluation (chargés d'étude ou membres de l'instance), ou du moins que ces préférences ont été suffisamment explicitées ou contrôlées pour que l'on puisse supposer qu'une autre évaluation répondant aux mêmes questions et utilisant les mêmes méthodes conduirait aux mêmes conclusions. Il s'agit notamment d'apprécier la rigueur et

l'honnêteté du travail de qualification (telle évolution, par exemple, est-elle "rapide", "normale" ou "lente" ?) et d'interprétation des données qui permet de passer de l'observation au jugement.

Possibilité de généralisation : il s'agit de savoir dans quelle mesure les conclusions portées sur les effets d'une action publique à partir d'observations faites sur un certain type de terrain pendant une période donnée peuvent être étendues à d'autres situations, à d'autres contextes, voire à des politiques différentes. Le plus souvent, les résultats d'une évaluation sont en effet utilisés pour éclairer des décisions relatives à des actions et à des contextes qui différeront en bien des points de ceux sur lesquels l'évaluation a porté.

Transparence : outre l'exigence d'un exposé complet et rigoureux des méthodes mises en œuvre, ce critère inclut l'idée que l'évaluation doit expliciter son propre "mode d'emploi" et ses limites : positionnement par rapport à d'autres évaluations possibles sur le même sujet, rappel des questions auxquelles il n'est pas ou incomplètement répondu, anticipation des objections possibles, etc . Cet effort de lucidité et de réflexivité est d'autant plus nécessaire que les évaluations sont généralement imparfaites, qu'elles laissent de nombreuses questions dans l'ombre, et que les résultats auxquels elles parviennent n'ont pas tous la même portée ni la même robustesse.